(12) UK Patent Application (19) GB (11) 2 280 827 (13) A

(43) Date of A Publication 08.02.1995

(21) Application No 9414078.7

(22) Date of Filing 12.07.1994

(30) Priority Data
(31) 933182 (32) 13.07.1993 (33) FI

(51) INT CL[6]
G10L 3/02

(52) UK CL (Edition N)
H4R RPCX
U1S S2105

(56) Documents Cited
GB 1528344 A US 5018136 A

(58) Field of Search
UK CL (Edition M) H4R RPCP RPCX RPV RPVA RPX
INT CL[5] G10L , H04B
Online databases: WPI, JAPIO, EDOC

(71) Applicant(s)
Nokia Mobile Phones Limited

(Incorporated in Finland)

P.O. Box 86, SF-24101 Salo, Finland

Nokia Telecommunications Oy

(Incorporated in Finland)

PO Box 44, Upseerinkatu 1, SF-02601 Espoo, Finland

(72) Inventor(s)
Ari Sinisalo

(74) Agent and/or Address for Service
T J Frain
Nokia Mobile Phones, Patent Department,
St Georges Court, St Georges Road, CAMBERLEY,
Surrey, GU15 3QZ, United Kingdom

(54) Speech compression and reconstruction

(57) The present invention relates to a method with which a speech signal can be compressed, respectively reconstructed, with the aid of two-dimensional algorithms developed for image processing. Both the compression and reconstruction can be thought of being formed from two different phases. In the first phase from a one-dimensional speech signal a two-dimensional image matrix (spectrogram) is formed in which the frequency components of a speech signal transformed from time domain to frequency domain with e.g. Fast Fourier Transform algorithm are presented frame by frame as a function of time. In the second phase of the method the image presented by the image matrix is compressed with a powerful two-dimensional compression algorithm, e.g. by JPEG. The reconstruction of speech is performed in inverse order; decompression (IJPEG) and inverse FFT (IFFT). The method is particularly appropriate for storing and transmitting speech when no strict real time properties are necessitated in the transmission or storing.

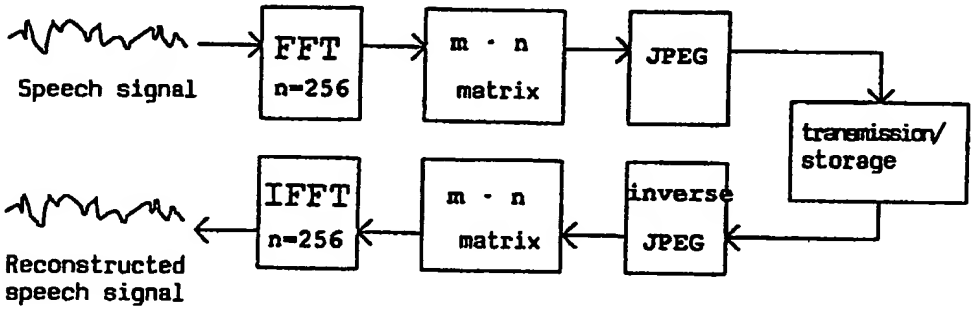

Fig. 2

GB 2 280 827 A

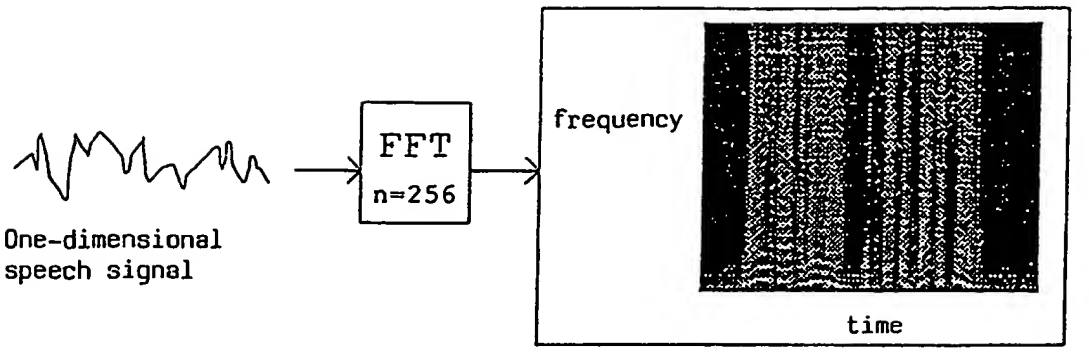

Fig. 1

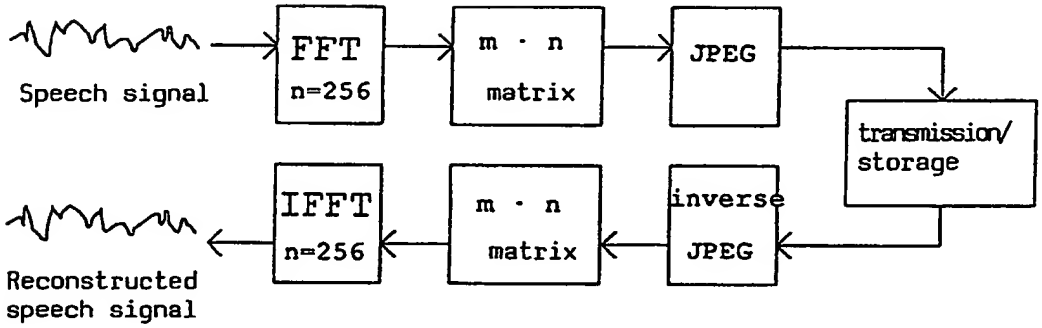

Fig. 2

File	Length	Size/Byte	Compression ratio	Bytes/s	Bits/s
orig.	6	48000	1.00	8000	64000
r_90	6	3182	15.08	530	4243
r_80	6	2077	23.11	346	2769
r_70	6	1631	29.43	272	2175
r_60	6	1365	35.16	228	1820

Fig. 3

Compression and reconstruction of speech signal

The present invention relates to a method for compressing and reconstructing a speech signal and a means for implementing the method.

The phrase digital signal compression is used to describe the packaging of a signal with the aid of a compression algorithm into a form in which the amount of the data to be processed is essentially reduced. Endeavours have been made to compress the amount of data corresponding to a signal in general e.g. into one tenth or one twentieth part of the original data amount. The compression algorithm is usually selected according to the purpose it serves. For the audio range signals, particularly for speech, specific compression algorithms of their own are provided, and the same goes for image processing.

Reconstruction of a signal refers in the present context to discharging of a compressed signal into the original form. Instead of reconstruction also term decompression is used. For decompression, the same algorithm is used as for compression, but in inverse order. Typically, the application of a compression -decompression algorithm is associated with signal transmission in a data communication channel, for instance in a telephone or data network, from a transmitter to a receiver. The signal is compressed before transmission and decompressed in the receiver end. The signal transmission has increased data density as more data can be included in one transmission bandwidth. A second general target for applying the compression-decompression algorithm is the storing of a signal. The signal is compressed when being stored e.g, on a diskette, and respectively, reconstructed when read in, in order to save the storing capacity. The practical compression algorithms are not without losses, with the exclusion of a few exceptions; therefore, a reconstructed signal is not completely identical with the original signal.

The compression ratio refers to the ratio of the amount of data of a non-compressed signal and the data amount of a compressed signal. The compression ratio is usually an optional variable of an algorithm, and it is determined to be appropriate for the purpose of the signal. The compression algorithms as such cause some losses compared with the original signal. An increase in the compression ratio will increase the amount of losses. For instance, in transmitting a speech signal can be used high compression ratios if merely intelligibility is selected as the criterion of quality, and particularly, if the insignificant amount of data obtained as an end result is emphasized. On the other hand, if also the maintaining of quality and tone of the sound are set for the criteria of quality, in addition to intelligibility, a small compression ratio is probably what we have to be content with as the details of the signal are no longer reproducible at high compression ratios.

Compression of a speech signal is in general based on coding a signal, i.e. an individual code is determined for different types of signal forms, in which the amount of data contained in a code is smaller than the amount of data in the original signal. Digital signal processing has made development of sophisticated speech coders possible, in which the interdependencies between samples of sampled speech are made use of, that is, short term prediction and long term prediction. The coding algorithms used in short term prediction are called Linear Predictive Coding, LPC, and such algorithms study the correlation between successive samples, whereas the algorithms used in Long Term Prediction study the long term correlation between successive base frequency segments. The long term prediction is applied in Regular Excitation-Long Term Prediction (RPE-LTP) coder wherein the sampling is carried out at 8 kHz frequency, and as a result of said coding, 20 ms frames containing 260 bits are obtained. In another significant coder, a code-excited linear prediction is used, also called stochaistic coding, a variation of the coding algorithm, a so-called Vector-Sum Excited Linear Excited Predictive

Coding, VSELP, in which a so-called codebook is used in accelerating the calculation.

The compression algorithms developed for speech signal are not very effective. The highest compression ratio of the algorithm achieved so far is about 10 to 15, i.e. the compressed signal is 1/10 to 1/15 of the data quantity of the original signal. In the development work of the present invention, endeavours have been made to provide a method with which the compression ratio of the speech signal can be made higher than what is possible in the prior art methods. The aim is achieved with the aid of the methods according to claims 1 and 3.

By applying the method of the invention the compression ratio of the speech signal can be at least doubled to the compression algorithms of the state of art, and not losing anything of the intelligibility and the characteristic features of the voice. The basic idea is to transform a one-dimensional speech signal frame by frame from a time domain to a frequency domain. According to a preferred embodiment, such transformation is performed with a Fast Fourier Transform algorithm, FFT, as it is the most common and easiest way of transforming the time domain to the frequency domain. A desired amount of successive frames is gathered from a speech signal, and a two-dimensional image matrix (like spectrogram) is produced in which the frames of the frequency domain are presented as a function of time. Reconstruction of speech is performed in inverse order using decompression and an inverse transformation of FFT.

By FFT, or a Fast Fourier transform, a discrete Fourier transform known in mathematics is implemented using approximation, whereby it has been possible to reduce the number of multiplications contained in said transform from n^2 to 2nlogn where n is the number of the calculation points. With the

aid of the Fourier transform a time domain signal can be described in frequency domain. A Fourier transformed signal includes both real and imaginary components. As an end result, an inverse transform of a Fourier-transformed signal produces the original signal. Today, the FFT transformer is a commercial component.

In a preferred embodiment, the number of Fourier transform points of a frame and the number of the frames to be transformed is the same, which with a view to the compression to be accomplished later is an optimal choice.

In a preferred embodiment, Joint Photographic Experts Group, JPEG-algorithm is used for the compression algorithm. The JPEG is a compression algorithm, standardized by ISO, for continuous toned still images. JPEG has been developed to compress colour and grey-level images, presenting so-called natural sceneries with slowly changing tones. The image matrix of the present invention produces a grey-level image appropriate for JPEG. An advantage of JPEG compared with other compression algorithms is that since it has been standardized, the means and circuit manufacturers are interested in manufacturing components implementing the algorithm.

In an advantageous embodiment, instead of the real and imaginary frames of the output of the FFT, equivalent amplitude and phase frames can be used for forming an image matrix, or also merely an amplitude frame.

A method and apparatus in accordance with the invention is described in detail by aid of the accompanying drawings, in which

Fig. 1 presents transforming of speech into image form with the aid of FFT transform

Fig. 2 presents a signal chain related to compression and reconstruction, and

Fig. 3 presents an example of applying the JPEG algorithm on different parameters.

The first phase of the compression method according to the invention concerns transforming one-dimensional speech signal to a two-dimensional image. The principle is presented in Fig. 1. A speech signal is fed to a FFT transformer, which in an example shown in the figure calculates a transformation for the 256 points of the frame of the speech signal, n = 256. The Fourier transformed real speech signal includes both a real and imaginary part. The image matrix can be produced by utilizing both the real part and the imaginary part, or respective amplitude and phase frames, or merely the amplitude frame. The two-dimensional presentation of Fig. 1 contains merely an amplitude frame. The horizontal axis illustrates time which consists of n-dimensioned successive frames in which a Fourier transform is accomplished. The vertical axis illustrates frequency variation at different points. The image is in practice an intensity image of the amplitude frames. Since a presentation of the signal frequency domain is always symmetrical in the range -1/n ... 1/n, the output frame of the FFT transformer can be halved, i.e. merely the positive part of the spectrum can thus be selected, that is, the one without defects.

In the second phase of the method the image is compressed using existing image compression algorithms. In test situations, with a standardized JPEG algorithm good compression ratios have been achieved, e.g. 20 to 30. JPEG is not without any losses, so that a given number of the details of the image are lost. In the signal chain presented in Fig. 2 different phases of the method can easily be detected. An original speech signal is conducted to the FFT transformer transforming a time domain signal frame to frequency domain. Prior to the actual transformation, the signal is also quantized by the FFT transformer, or the quantization may also be performed with a separate

quantization circuit. An image matrix is gathered from the frequency domain frames, the dimensions thereof being m, i.e. the number of frames in time domain, and n, the number of frequencies equivalent to the time domain frames. In practice, the image matrix corresponds to a sample of a few seconds of a speech signal. With the FFT transformer of Fig. 2, transformation is calculated for the 256 points of the frame, n = 256. It is preferred to select for the number of frames to be gathered the same as the length of the frame, thus, m = 256. The compression, JPEG, is carried out before transmitting or storing the signal, said operation being illustrated by block 'transmission/block'. The decompression IJPEG, i.e. Inverse JPEG, is performed in association with reception /read operation of the signal. Decompression yields in principle a similar image matrix as at the first end of the signal chain, this being, however, somewhat transformed in details because of the losses caused by the compression and transmission. In test situations, the greatest losses occurred in the transmission path of the signal, while the losses caused by the compression itself were small. The image matrix is transformed back into an one-dimensional signal with the aid of an inverse FFT transformation, IFFT.

The present method thus disclosed has been tried with the JPEG algorithm at various compression ratios using both an image assembled from the amplitude frames and separate images assembled both from the real and the imaginary frames. The use of an amplitude frame alone simplifies the calculation. On the other hand, omitting the phase frame generates distinct disturbances in the reconstructed signal. Using both real and imaginary images yields the best qualitative end result at the reconstruction phase, but the compression ratio remains below five without losing the identifiability of the speech. By using only the amplitude frames will give a good compression ratio 20:1 to 30:1 and the identifiability of the speech is maintained. Although the JPEG algorithm causes losses in the frequency domain signal,

their impact on the time domain speech signal is surprisingly low.

In an experiment carried out it was proved that transmission of a signal without compression caused more disturbances in the signal compared with additional disturbances brought about compression. By enhancing the algorithms, a greater amount of disturbances caused by the compression can be reduced.

In the table of Fig. 3 test results are presented concerning the compression of an eight bit speech signal of the length of six seconds when JPEG algorithm is used. For the JPEG algorithm, used as a parameter, a desired permanence of the details of the image can be provided. The less details are preserved, the higher will be the compression ratio. The figure in the file name of the table, e.g. r_90, illustrates the preservation degree of the details as per cents. At r_60, the speech is still clearly intelligible and the speaker can be identified. The other columns in the table illustrate the amount of data corresponding to the signal (Size/Byte, Bytes/s, Bits/s) at various compression ratios. The compression algorithm is not restricted by the method according to claim 1 to the exemplary JPEG algorithm described above.

By the method of the present invention, a speech signal can be compressed even to 1/30 part of the amount of data provided by the original signal by utilizing the existing methods known in the art, and by maintaining the identifiability of the speech and the speaker. The method is advantageous particularly in transmitting speech and in storing speech. By means of calculation, it is found that speech lasting one hour can easily be stored in a standard 1.44 Mbit disk. A result thus obtained represents at least a doubled amount of data compared with the one-dimensional compression algorithms developed for the speech signal. In addition, the characteristic features of the speech can be maintained by the aid of the method, even at high

compression ratios. A drawback of the method is that it is heavy as regards the calculation, with the inclusion of a great number of phases. Relief thereto can be provided through the development of circuit technology. The method is not appropriate for real-time processing, either, because speech signals have to be gathered for several seconds prior to the actual compression, in order to assemble an image matrix.

Targets for the method of the invention are particularly the speech storing and sound mail applications, which do not require strict real time properties. As a target to be adopted in the future, multimedia programs of a microcomputer are conceivable, in which the transmission and storing of speech form an essential part. Currently, the application of the method is facilitated by both the FFT transformers implemented with ASIC circuits and by the image processing algorithms.

Claims

1. A method for compressing a speech signal, wherein
- successive frames of a desired length are separated in the time domain from a speech signal,
- the time-domain frames are conducted to a time domain/frequency domain transformer, from the output of which the frames are received in frequency domain,
- the frequency-domain frames are arranged into an image matrix, (spectrogram), in which the frequency-domain frames are presented as a function of time,
- the image matrix is conducted to a means implementing the compression algorithm of said image,
- from the output of which a compressed speech signal is provided.

2. Method according to claim 1, wherein the time domain/frequency domain transformer is a Fast Fourier Transformer.

3. Method for reconstructing a speech signal compressed with the method according to claim 1, wherein - a compressed speech signal is conducted to a means implementing a decompression algorithm, the output whereof being the image matrix presenting the frequency-domain frames as a function of time,
- from the image matrix, the frequency domain frames are separated, said frames being conducted to a transformer performing the frequency domain/ time domain transformation, for the output of said transformer the frames forming the time domain speech signal are obtained.

4. Method according to claim 3, wherein the frequency domain/time domain transformer is a transformer performing Inverse Fourier Transform.

5. Method according to claim 1 or 2, wherein the compression algorithm is a Joint Photographic Experts Group, JPEG, known in itself in the art, and the decompression algorithm is an inverse JPEG, IJPEG, known in itself in the art.

6. Method according to claim 2 or 4, wherein the frequency domain frame consists of separate real and imaginary frames.

7. Method according to claim 2 or 4, wherein the frequency domain frame consists of separate amplitude and phase frames.

8. Method according to claim 2 or 4, wherein the frequency domain frame consists merely of amplitude frames.

9. Method according to claim 1 or 2, wherein one half of a frequency domain frame is used for forming a two-dimensional image matrix.

10. Method according to claim 2 or 4, wherein the number of the Fourier transform points of a frame and the number of the frames to be transformed is the same, so that a square image matrix is obtained.

11. A means for compressing a speech signal, wherein the means comprises
- an A/D converter for converting a speech signal to digital form,
- first means for dividing the speech signal into consecutive frames,
- a first transformer to implement the time-domain/frequency-domain transform,
- second means for producing a two-dimensional image matrix from the output signals of the first transformer,
- third means for implementing a compression algorithm on the image matrix, and
- memory space for storing the intermediate and end results.

12. Means according to claim 11, wherein the first transformer is a transformer implementing a Fast discrete Fourier Transform.

13. Means for reconstructing a compressed signal with a means according to claim 9, wherein the means comprises

- fourth means for implementing a decompression algorithm,
- fifth means for separating the frequency domain frames from a two-dimensional image matrix,
- a second transformer to implement frequency domain / time domain transform,
- sixth means for connecting in succession the time domain frames,
- a D/A converter for converting a digital speech signal to analog form, and
- memory space for storing the intermediate results.

14. Means according to claim 13, wherein the second transformer is a transformer performing Fast inverse Fourier Transform.

15. Means according to claim 11 or 13, wherein the third means are the means implementing the JPEG algorithm and the fourth means are the means implementing the IJPEG algorithm.

16. A method for compressing a speech signal substantially as hereinbefore described and with reference to the drawings.

Patents Act 1977
Examiner's report to the Comptroller under Section 17
(The Search report)

12

Application number
GB 9414078.7

Relevant Technical Fields

(i) UK Cl (Ed.M) H4R: RPCP, RPCX, RPV, RPVA, RPX

(ii) Int Cl (Ed.5) G10L; H04B

Search Examiner
MR S SATKURUNATH

Date of completion of Search
25 OCTOBER 1994

Databases (see below)
(i) UK Patent Office collections of GB, EP, WO and US patent specifications.

(ii) ONLINE DATABASES : WPI, JAPIO, EDOC

Documents considered relevant following a search in respect of Claims :-
1, 2, 5-12, 15, 16

Categories of documents

X: Document indicating lack of novelty or of inventive step.

Y: Document indicating lack of inventive step if combined with one or more other documents of the same category.

A: Document indicating technological background and/or state of the art.

P: Document published on or after the declared priority date but before the filing date of the present application.

E: Patent document published on or after, but with priority date earlier than, the filing date of the present application.

&: Member of the same patent family; corresponding document.

Category	Identity of document and relevant passages		Relevant to claim(s)
A	GB 1528344	(GEN) - see especially Figure 2	1, 11
A	US 5018136	(GOLLUB) - see lines 66-69, column 2, lines 1-10, column 3, columns 8-10	1, 11

Databases: The UK Patent Office database comprises classified collections of GB, EP, WO and US patent specifications as outlined periodically in the Official Journal (Patents). The on-line databases considered for search are also listed periodically in the Official Journal (Patents).